Supplemental Table 1 Description Of Studies' Findings

Guideline	First Author Year	Measure of Quality	Results Mean (SD) / % Count	Authors' Conclusion
CONSORT	Adie, S 2013[1]	22 item-CONSORT checklist expressed 2001 version	The mean adjusted CONSORT score was 12.2 (SD = 3.8) out of 22 items	Existing studies assessing the reporting of randomized trials in surgery are **deficient** and **not adequately reported**. Trial authors need to be made aware of existing guidelines for reporting and journal editors should insist and assess compliance with these guidelines.
	Agha, R 2016[4] (combined STROBE, CONSORT and PRISMA guideline)	22 item-CONSORT checklist expressed 2001 version	No RCTs were found for period 1; Median compliance for the 3 studies found for period 2 was 50% (range from 36-50%); Median compliance for the 10 studies found for period 3 was 70% (range from 54-82%)	The overall guideline compliance following implementation of the policy **increased** for RCTs by 40% (ranges from 50-70%). The compliance of the guidelines can be increased by implementing a policy mandating the submission of a completed CONSORT checklist for RCTs.
	Al-Namankany, A 2009 [8]	Modified 34-item CONSORT checklist	Compliance varied across items and articles. Good compliance of articles to CONSORT for introduction sections (96-98%), discussion sections (96-98%). Poor reporting in randomization methods (5-9%), description of sample size calculation (4%), intention-to-treat analysis (1%).	Quality of reporting of RCTs in paediatric dental journals was **generally poor**, with negligible improvement after the publication of CONSORT statement.
	Alvarez, F 2006[9]	Criteria originating from the CONSORT checklist	11% of RCTs displayed a good methodological score in 1997 compared to 28% in 2006 with a statistical significance of p=0.03.	The quality of RCT reporting remains **suboptimal** even after a revision of the CONSORT statement. More than 50% of RCTs published in 2006 did not report power calculation or randomization method despite CONSORT adoption and improvement in reporting quality in both the JAAD and the BJD.
	Anttila, H	Modified 33-item	48% (almost half) of the	No clear difference showed

	2006[10]	2001 CONSORT checklist	applicable items were reported adequately.	in the quality of reporting between 1990 and 1997 and 1998 and 2002. **Poor** reporting is found in the study. The improvement for the quality of reporting of trials in clinical implications is clearly needed. Authors should be further encouraged to follow the CONSORT criteria when reporting.
	Areia, M 2010 [11]	Application of CONSORT/STARD	15.7 (2.2)	Level of adherence is **medium** for quality of reporting in diagnostic endoscopy.
	Augestad, K 2012 [12]	CONSORT adherence, Jadad	30.75 (4), 40% of the trials had a Jadad score of ≥3 points.	Level of adherence is **low** for quality of reporting for RCTs of disease specific clinical decision support.
	Balasubramanian, S 2006 [13]	Modified CONSORT score, Allocation concealment as assessed by Schulz et al, Jadad score	Medians of the modified CONSORT score were 85.45 (81.09-86.13) and 68.97 (62.89-73.11) for RCTs from medical and surgical journals, respectively, 13% clearly explained allocation concealment, 37.7% of RCTs had a Jadad score of ≥3.	Quality of reporting of surgical RCTs was **suboptimal**, and reporting in surgical journals was inferior to surgical trials in medical journals. "We found that the quality of reporting of general surgical RCTs leaves considerable room for improvement".
	Bath, F 1998 [14]	33 criteria of the CONSORT statement and 53 additional factors relevant to acute stroke or trials in general. Trial quality was also assessed with a 7-point scale.	Median total report quality was 40/86 (range 15-61). Median CONSORT criterion was 19/33 (9-29).	**Poor** quality for acute stroke RCTs. "We believe that authors should follow the CONSORT guidelines and that referees and editors should ensure this happens".
	Bian, Z 2006 [16]	63-item revised CONSORT checklist designed for Chinese Herbal Medicine clinical studies	Median score of overall reporting quality was 32% (8%)	Overall quality of reporting of Chinese herbal medicine (CHM) RCTs was **poor**. Need to improve reporting in clinical trials in this area. "To improve the quality of reporting of RCTs of CHM, we recommend adopting a revised CONSORT checklist that includes items specific

				to CHM. We also recommend that editors of CHM journals require authors to use a structured approach to presenting their trials as a condition of publication".
	Borg Debono, V 2012[18]	15-items and 3 methodologically related items from 2010 CONSORT checklist	Only 4 out of 15 items being reported in over 90% of the articles. Less than 50% of the articles reporting any of the 3 key methodological items	Overall quality of reporting of the RCTs assessed was **poor to moderate**. Reporting of the key methodological items was **poor** with less than 50% of the articles reporting any of the 3 key methodological items. Enforcing the use of the CONSORT statement by requiring authors to submit a CONSORT checklist is greatly helpful for improving the quality of reporting.
	Bousquet, P 2010 [19]	Reporting of procedure, randomization, dropouts, strict conduct of intention-to-treat analysis, sample size calculation, which was assessed with 8 items of the CONSORT statement	4/94 studies met the 8 items of the CONSORT statement criteria	RCTs in subcutaneous immunotherapy and sublingual immunotherapy had **poor** reporting quality. Encourage more use of CONSORT statement.
	Cairo, F 2012[21]	Adherence to modified 35-item 2010 CONSORT checklist	9% of articles showed adherence to CONSORT statement	Quality of reporting was only **partially** improved over time and there is **lack** of CONSORT adherence
	Capili, B 2010 [22]	Adherence to CONSORT/presence of harms guidelines in RCTs	17(range : 14-21) for CONSORT; 7/22 = 77.3% for harm	Level of adherence is **bad** for quality of reporting for RCTs on acupuncture for pain reduction.
	Cavadas, V 2011 [23]	25-item 2010 revised CONSORT statement	No combined data: only a few items were reported in less than 50% of the studies; some items were reported in more than 90% of the studies.	"RCTs in [pelvic organ prolapsed] are scarce. The quality of reporting is **suboptimal** in many aspects and has not improved in recent years".

	Choi, J 2014[24]	Adherence to the 38-item CONSORT statement for non-pharmacological trials (CONSORT-NPT)	Average of 11.3 CONSORT items (29.6%). Among the 28 included articles, the reporting percentage in each of the articles was 21.6 to 56.8%.	The results demonstrated a very **low** reporting quality. The endorsement of reporting guidelines is limited in traditional medicine journals in Korea since many items in research studies were **far from satisfactory**.
	Chowers, M 2009 [25]	CONSORT guidelines for adverse events, adjusted to the design of the HAART trial.	No combined score: harms were reported in only 24% of trials, 1/49 reported on adverse events collection method.	Large variability and a **lack of standard reporting** of adverse events between trials, many trials did not adhere to CONSORT recommendations.
	Daitch, V 2016[27]	Modified 2010 CONSORT checklist with additional dimensions relevant to patients with cystic fibrosis.	56.9% of articles did not define a primary outcome; 70.6% did not provide details on sample size calculation; only 31.4% reported on the subgroup or separated between important subgroups.	Most studies were methodologically **weak**. Poor quality studies included studies with a small number of patients but without sample size calculation, a relatively short-term intervention, or without examining the outcomes that are important to the patient. Improvement over the years has been minor.
	Dasi F, 2012[28]	46-item derived from the 2010 CONSORT checklist	Median equals 30 with a range of 20 to 42.	The overall quality of the articles **varied substantially** among the publications with scores ranged from 20 to 42 which indicated some aspects of the articles must be improved.
	de Vries T, 2010[31]	Adequate reporting of adverse drug reactions	Mean of 3 and 18% of articles scored 6 or higher.	**Insufficient** reporting quality in adverse event reporting in RCTs of children.
	Delaney M, 2010[29] (combined STROBE and CONSORT guideline)	Adherence to CONSORT and STROBE checklists	RCTs (CONSORT): 43% failed to describe study deviations and sources of potential bias; 46% reported on the method used to generate the random allocation sequence; 11.3% evaluated the success of blinding.	Both observational studies and RCTs **missed reporting some important factors**. There were gaps in methodologic reporting. Using CONSORT and STROBE checklists may improve the **deficiencies** in reporting.
	DeMauro S, 2011[30]	The presence or absence of 11 quality criteria	The median number of criteria fulfilled by each article was 9 (IQR: 8-10).	The quality of reporting of infant and neonatal RCTs is **inconsistent**, particularly in

		from the 2010 CONSORT checklist	The proportions of studies that fulfilled each quality criterion ranged from 50% to 99%.	the pediatric journals.
	Dias S, 2006[32]	Cochrane Handbook (Higgins and Green, 2005) and the the CONSORT statement	24 (15%) were found not to be randomized, despite claims. 10 (6%) provided adequate details on the methods of randomization and allocation concealment. 3 (2%) had sufficient details extractable to allow for an intention-to-treat analysis of the outcome 'live birth'.	Even though there showed some improvement in some subfertility-specific issues, the quality of reporting of RCTs still **needs to be improved**.
	Ethgen M, 2009[33]	CLEAR NPT – a checklist to evaluate RCTs of non-pharmacological treatments	Most studies failed to report 8/12 quality indexes in the checklist. Reporting of generation of allocation sequence was adequate in 38.8% of studies, treatment allocation in 26.3%, intention to treat analysis in 70.0%.	**Inadequate** reporting amongst trials involving stents. "The current reporting of results of RCTs testing stents needs to be improved to allow readers to appraise the risk of bias and the applicability of the results".
	Eyawo O, 2008[34]	Revised CONSORT checklist to assess reporting of each items on the checklist in counts (percentage)	14/16 items ranged from 2-47%, the other 2 items, sample size determination and reporting of masking were reported in 72% and 75% of the articles.	**Deficiencies** in the design, planning, and reporting of non-inferiority and equivalence trials in ophthalmology literature.
	Fan F, 2014[35]	Modified 37-item (25 primary and 12 secondary) from the 2010 CONSORT checklist	Mean number of 37 reporting items for the 21 included articles was 17. 11.0 (52.4%) reported the title and abstract; 21.0 (100.0%) reported the introduction section; 9.5 (45.4%) reported the methods section; 10.8 (57.1%) reported the results section; 16.7 (79.4%) reported the discussion section; 3.7 (17.5%) reported the other information section.	Reporting of RCTs were **not consistent** with the 2010 CONSORT statement. Adherence to the CONSORT statement was effective in improving the reporting of RCTs.
	Farrokhyar F, 2007[36]	Modified CONSORT statement and	51.7 out of 105 (11.5)	The total reporting quality of trials in this review varied substantially between

		added factor relevant to surgical trials and CABG surgery		publication (35-96 out of a possible max score of 105) The results showed that there is a **need for improvement in quality of reporting.**
	Froud R, 2012[41]	Number and Percentage of studies satisfying the revised 11 item consort checklist	Most items were reported in adequate percent of studies, 5/11 reported in 78-100% of the studies.	Their results suggest that cluster randomized trials in oral health are of reasonable quality with respect to the key criteria of accounting for clustering in the design and analysis.
	Fung A, 2009[42] (combined STROBE and CONSORT guideline)	Presence or absence of CONSORT (Max score of 37 points) statement indicators	CONSORT: median and mean values of 89% and 83%, respectively.	Overall level of reporting is acceptable, and has improved since the creation of CONSORT and STROBE.
	Gagnier J, 2006[43]	Mean CONSORT score based on 42 items and the percentage of items reported	18.92 out of 42 (5.54) and 45% of items were reported across all trials	"We found that reports of RCTs of herbal medicine interventions reported less than half of the necessary information in their published results". **Overall adherence is low.**
	Gao J, 2015[44]	An overall quality score with 15 items from the revised CONSORT 2010 statement	The mean overall quality score was 7.10 with a standard deviation of 1.95. The level of reporting was below average.	The general level of reporting was not high. "Randomization" and "harms" were the lowest ratings. The reporting quality of RCTs in recurrent miscarriage is **unsatisfactory**. The results stress the need to improve the reporting quality of RCTs on recurrent spontaneous abortion in China because of the crucial methodological issues of allocation concealment, blinding and sample size calculation.
	Gohari F, 2016[46]	Percentages for included articles which reported each item from the 37-item CONSORT 2010 statement	The percentages varied from 0.5% to 98.4%. 3b and 7b were the lowest and 2b was the highest. Mean percentage for the single item from the 37-item was 43.2%.	The quality of reporting of RCTs on diabetes in Iran was **suboptimal** and **incomplete**, especially in the methods section. Adherence to CONSORT guidelines seems **deficient**.

	Halpern S, 2004[48]	Percentage of articles that reported each applicable item of the modified CONSORT checklist and count of articles complying with modified CONSORT items	In the 23 articles in Anesthesia and Analgesia, the median percentage of correct CONSORT items was 63%.	**Poor** – Total number of items that are inadequately reported is high in current RCT literature with obstetric anesthesia.
	Herdan A, 2011[50]	22 item CONSORT checklist expressed	On average 87.4% of the CONSORT items were reported.	The reporting quality has improved significantly in the period after dissemination of the CONSORT statement, however reporting of adverse events **needs attention**.
	Huang D, 2015[51]	Adherence on individual items of 22-item 2010 CONSORT statement	The percentages for each item varied from 2.5% to 100.0%.	The quality of reporting of RCTs for effect of laparoscopic and open surgery for colorectal cancer in China was **poor**. Reporting of methodological items such as randomization, blinding, and intention-to-treat analysis were **deficient**.
	Hui D, 2012[52]	37-item 2010 CONSORT checklist as well as reports on key methodologic index and Jadad score	Median CONSORT score was 9 with IQR from 7 to 11. < 50% of studies reported key areas included trial design, sample size calculation, details regarding to randomization and blinding, outcomes with appropriate statistics and limitations of research design.	Overall reporting quality of CONSORT score, key methodologic index or Jadad score was **low**. Multiple areas of reporting quality for supportive and palliative oncology studies were **deficient**.
	Karpouzis F, 2016[54]	CONSORT modified by authors - 22 items from CONSORT 2010 (excluded items 21, 22, and 24) and 9 items from CONSORT for non-pharamcological treatments statement (included 1, 3, 4a,	The overall quality of reporting score, ranged between 10 and 33 with median score of 26.0 (IQR = 8.00).	Univariate analysis: journal type, industry funding, positive finding

		4b, 4c, 8, 13, 15, and 'new item'). See Table 2 and 3.		
	Kiehna EN, 2011[55]	CONSORT (max. score 44) JADAD out of 5	26.4 out of 44 (range: 17–38)/67% of studies had no description or the prestudy sample size calculation, 63% did not describe whether subjects, treatment providers or assessors/analysts were blinded	The quality of reporting of RCTs in neurosurgical journals remains suboptimal
	Kim KH, 2016[56]	CONSORT 2001 for parallel RCTs and STRICTA for Korean RCTs	Early period (1996-2004): Mean=9.5 (8.9-10.2) Late period (2005-2011): Mean=10.6 (10.2-11.1) Difference: p=0.0082 Of 103 RCTs published in the late period, there was considerable incompleteness of reporting in items related to the study design, implementation, reporting and interpretation. Individual items reported in figure 2.	Overall, the completeness of reporting of Korean RCTs of acupuncture was suboptimal, which could represent a significant obstacle to the establishment of a sound evidence base.
	Kober T, 2006[57]	CONSORT based on 14-item criteria	75% of studies reported only six of the 13 items; only 14% reported randomization process; only 13% provided details about concealment of allocation; only 13% provided a statement on study power; only 12% used intention-to-treat analysis	Articles of Hodgkin's lymphoma published after 1996 do not conform to the CONSORT recommendations
	Ladd BO, 2010[58]	Assessment of 36 of the items from the CONSORT Statement based on a score out of 36	24.43 out of 36 (3.27)	The overall level of adherence to CONSORT has improved since 1994, and continues to remain highest among studies that have been published within journals that have adopted the CONSORT guidelines
	Lee SY, 2013[59]	CONSORT 2010	The mean (SD) CONSORT score was 11.2 (3.36) of 23	Number of authors, number of trial participants,

			items (48%; range, 3.83-18.17).	multicenter vs single-center studies
	Li JL, 2014[62]	CONSORT 2010 and JADAD (5 point scale)	See Table 2 for individual items. See Figure 7 for illustration of mean scores of reports before and after 2010. Scores ranges from 1-24, with most of them within the range of 4-11.	Not reported
	Li JY, 2011[61]	Score out of 40 based on a 40-item modified checklist based on the CONSORT Statement	42% of the studies included explained how sample size was determined; 14% of studies described whether or not outcome assessors were blinded	The reporting quality of these trials is suboptimal and substantial improvement is required to meet the CONSORT guidelines. Almost 50% of the trials we reviewed did not satisfy more than half of the criteria in the modified CONSORT checklist, and only 23% of RCTs provided adequate details of Tai Chi intervention used in the trials
	Liu LQ, 2013[65]	CONSORT 2010 and JADAD	On average, 14 out of 30 CONSORT items (range 1–26) were addressed in each trial report.	Journal impact factor, commercial funding, no funding
	Liu XT, 2015[66]	CONSORT 2010 and JADAD	JCIM (adopter): Mean=23.18 (SD=7.92) CJIM (nonadopter): Mean=15.61 (SD=4.58)	Not reported
	Lu J, 2015[69]	CONSORT 2010 checklist (study did not evaluate all 25 items on checklist)	Among all CONSORT checklist items examined, the improvement in reporting the methods of randomization sequence generation and allocation concealment noticed in this study is of great importance, as studies have shown that it is strongly associated with effect estimates. For any RCT, the method of randomization is a key component to minimize any measured and unmeasured differences between the comparison	Below acceptable levels

			groups.	
	Lu L, 2011[70]	Percentage of articles that reported each applicable item of the CONSORT checklist	Sample size: only one (2.2%) of the papers mentioned sample size calculation. Randomization: 12 studies (26.1%) were deemed to have authentic randomization. Blinding: 36 papers (78.3%) provided no information about blinding of either participants or investigators. Reporting of baseline characteristics: 39 papers (84.8%) reported the details of the baseline characteristics of participants. Length of follow-up: 22 papers (47.8%). There was no information provided on the length of time for which participants were followed. Loss-to-follow-up: a total of 36 studies (78.3%) failed to report dropout rates. Statistical reporting: only one paper (2.2%) did not report what statistical methods they had used	Findings indicate that the reporting quality of RCTs needs improvement for RCTs on the treatment of cancer pain in People's Republic of China
	Marshman Z, 2010[73]	56 criteria based on the CONSORT Statement	27/56, with variation between journals (23.2 to 27.7)	Poor adherence to the CONSORT checklist in RCTs in dental health

	McCormick F, 2013[74]	2010 CONSORT checklist criteria (each item on the 37-point checklist was evaluated and given a score: 0 for no description or criterion not met; 1 for inadequate description, and 2 for adequate satisfaction of criterion) and JADAD	The mean CONSORT Criteria Score was 70% (range, 30-98%, SD 16). One in 5 studies (20%) satisfied 75% or more of the criteria. Deficiencies commonly identified in the majority of papers include: lack of randomization type description; lack of study funding source disclosure; trial registration; and full protocol general access. Important, a power analysis was reported in only 35 out of 54 studies (64.8%), and sufficient blinding was absent in 40% (21/54)	High quality in past decade
	Moberg-Mogren E, 2006[76]	Average NMECI score (0–201 subitems scale)	104.2 (32.9)	Less than half of the articles met criteria of these subitems in selected RCTs relevant to occupational therapy
	Moher D, 2002[77]	CONSORT checklist, frequency of unclear allocation concealment, and a five-point quality assessment instrument (Jadad)	12.7/32 of the CONSORT checklist included; 81.3% unclear allocation concealment; 1.9/5 for the Jadad assessment scale	Overall, there was no difference in the PedCAM RCTs and conventional medicine quality, with both types achieving 43% of their maximum possible outcome
	Montané, E. (2010)[78]	Revised CONSORT checklist, 22 items	10.5 (2.7)	Quality was good in 23 (25%) of the articles and poor in 69 (75%) of the reports for RCTs on the efficacy of analgesic drugs in postoperative pain after TOS
	Montgomery , A.A. (2011)[79]	Qualitative look	N/A	Varying level of reporting quality factorial trials of complex interventions in community settings
	Norton-Mabus, J.C. (2008)[81]	NMNECI (212 subitems)	119.5 (25.48)	Article consistency with CONSORT Statement was less than 60%. Occupational therapy RCT had higher consistency with the

				instrument, scoring higher than articles in speech therapy
	Ntala, C. (2013)[82]	CONSORT 2010 statement with non-pharmacological, non-inferiority and equivalence, cluster and pragmatic extensions	4 trials adequately reported <50% of the items, 15 trials adequately reported 50–60% of items, and 16 adequately reported >60% of items.	Overall QOR was **suboptimal**. Access to the full trial protocol was the least well reported item, present in only approximately one in five trials.
	Parsons, N.R. (2011)[84]	Overall compliance calculated as the weighted mean of the compliance rates for the seven selected journals, using a previously made questionnaire	59% (CONSORT)	Very few papers fulfilling all criteria; general lack of statistical rigor
	Patel, M.X. (2015)[85]	CONSORT for superiority and non-inferiority/equivalence designs	Mean percentage for reporting on methodology for some items of CONSORT guidelines was 55.1%. Mean percentage for reporting on outcomes and analyses was 48.8%.	Overall QOR was **suboptimal** in phase II and III trials for more recently developed and/or licensed antipsychotics.
	Péron, J. (2012)[87]	CONSORT (2001 and 2010)	Mean 2001 CONSORT OQS for all items was 13.4 (ranged from 6-18, 95% CI =9-17). Mean 2010 OQS was 19.3.	Overall QOR **significantly improved** overtime, with some areas remaining **poor**.
	Piggott, M. (2004)[86]	Compared RCTs of three different time period cohorts, with the CONSORT (condensed, 13-item) checklist	Quality of reporting variable; 30% of trials or less used true randomization, allocation concealment, intention-to-treat analysis, and power calculations	Quality of reporting over time cohorts was variable, no consistent improvement over time. Quality of reporting remains poor for RCTs in specialized palliative care literature
	Plint, A.C. (2006)[89]	22-item checklist from the CONSORT Statement	Standardized mean difference between CONSORT-adopting journals and nonadopters was 0.83 (95% CI, 0.46–1.19)	Journal adoption of CONSORT is associated with improved reporting of RCTs

	Prady, S.L. (2008)[90]	5 categories from CONSORT	Pre-CONSORT (1994-1995) reporting was 23.2% (95% CI 16.8-29.6%); Post-CONSORT (1999-2000) reporting was 33.8% (95% CI 27.8-39.8%); Post-revised CONSORT (2004-2005) reporting was 51.0% (95% CI 45.4-56.5%).	Overall QOR **significantly improved** overtime.
	Pratoomsoot, C. (2015)[91]	22-item elaborated CONSORT for RCTs of herbal interventions	Percentages for each item were reported and ranged from 2% to 68%.	Overall QOR was **poor** for some items.
	Rikos, D. (2016)[95]	Revised CONSORT 2010	The average CONSORT compliance score was 68.2% (23.7–94.7%).	Overall QOR was **suboptimal**. Only 20 of the 38 items of the checklist (52.6%) were addressed in 75% or more of the studies published in the period between 2000 and 2015.
	Rios, L.P. (2008)[94]	Overall quality score, which is a 15 point overall reporting quality score made from CONSORT checklist	10 (2.03)	Suboptimal reporting quality in an endocrine journal
	Scott, P. (2012)[97]	27 items based on CONSORT 2010	The percentages for included studies ranged from 7% to 77%.	Overall QOR was **incomplete**.
	Stevely, A (2015)[101]	CONSORT 2010	The median proportions (IQR) of RCTs meeting complete and at least partial compliance in reporting criteria of checklist items was 81% (53% to 91%) and 93% (78% to 97%), and a minimum of 12% and 22%, respectively.	Suboptimal
	Strech, D. (2011)[102]	A checklist based on the CONSORT Statement	There are 72 items on the checklist; 42% were reported adequately and 25% were reported inadequately	While some trial-related information is well reported, a good part of the reporting quality of RCTs in bipolar disorder falls well below the required and practically feasible level for many aspects essential for the

				adequate interpretation of methodological quality and clinical relevance. Authors should be further encouraged to follow the CONSORT criteria. No consistent trend could be shown for improvement in the quality of reporting over time, or for reporting essential methodological items differently. There is a consistent trend toward better reporting in journals that endorse the URM
	Thabane, L. (2007)[103]	Percentage of studies satisfying each of the 44 CONSORT criteria	26.25 (4.51) and 60% adherence for reporting criteria: 90% satisfied criteria for the introduction; 19% for the methods; 75% the study protocol, 70% for the results	Overall, the quality of reporting is suboptimal in RCTs of weight loss intervention. Key reporting criteria that may impact the validity and generalizability of the results were adequately reported
	Turner L. (2012)[106]	CONSORT statement (22 items of the CONSORT 2001 checklist, plus four items relating to the reporting of blinding, and one item of aggregate CONSORT scores.)	None	Suboptimal
	Wang G., (2007)[109]	CONSORT	The mean (SD) number of the modified CONSORT checklist items reported across all trials was 11.82 (5.78), or 39.4% of the 30 items	Poor
	Wang, P. (2013)[110]	37-item modified CONSORT	The average reporting percentage was 45.0%. In the Journal of Traditional Chinese Medicine, Chinese Journal of Integrated Traditional & Western Medicine, and the China Journal of Chinese Materia Medica, the	Overall QOR was **insufficient**. Suboptimal reporting of bias correction methods could potentially imply most GS trials stopping early are giving biased results of treatment effects

			average reporting percentage was 42.2%, 56.8%, and 46.0%, respectively.	
	Weingärtner, V. (2016)[112]	CONSORT PRO	On average, 4.4 (SD 2.5) of the 14 CONSORT items were met.	Overall QOR was **inadequate**, despite some improvements.
	Walleser, S 2011[108]	Consolidated standard of reporting trial-CRT (CONSORT-CRT)	34% inadequately reporting on more than half of the CONSORT-CRT criteria.	The quality of reporting in CRTs **needs improvement**. This will hopefully improve implementation and planning.
	Wangge, G 2010[111]	Extension of the CONSORT statement for NI and equivalence trial.	No blinding in 34.0%, non-inferiority margin in 97.8% with only 45.7% reporting method of determining the margin.	Adherence improved slightly after CONSORT for non-inferiority trials.
	Yao, A.C. (2014)[116]	23-item 2008 CONSORT NPT extension	The mean CONSORT score of the 65 RCTs was 8.9 out of 23 (39%, range 3.0–14.7, SD 2.49).	Overall QOR was **low**.
	Zhong, Y 2011[121]	Number of studies describing each of the 38 modified consort items	Of the 38 CONSORT items, only 5 items were described in more than 80% of the 153 included.	Adherence was **suboptimal** for two-group parallel randomized controlled clinical trials of multi-herb formulae.
	Zhao X, 2016[119]	37-item 2010 CONSORT checklist	The reporting percentage in each of the 68 articles ranged from 24.3% to 73%. A total of 21 (30.9%) articles reported more than 50% of the items.	The reporting quality of placebo-controlled RCTs on the treatment of diabetes with traditional Chinese medicine had **improved** after the publication of 2010 CONSORT statement.
	Zheng S, 2016[120]	25-item 2010 CONSORT checklist	The mean article CONSORT score was 55.4% (ranged from 23.3–93.8%, SD 17.2%).	The reporting quality of RCTs in in heart failure with preserved ejection fraction was **inadequate** even though the CONSORT score increased over time.
	Zintzaras E, 2010[122]	17 item CONSORT checklist	17 CONSORT checklist items were reported in 7/18 studies and 9/17 CONSORT checklist items were reported in all 18/18 studies.	Proper assessment of the credibility and generalizability of the results can be ensured by reporting quality.
	Ziogas 2009[124]	24-item questionnaire based on the	75% of the studies addressed 13 out of the 24 items of CONSORT	Reporting on myeloid malignancies remains **unsatisfactory** and requires

		CONSORT checklist.	statement.	further improvement to properly assess the validity of clinical research.
PRISMA	Adie, S 2015[2]	27-item PRISMA and 11-item AMSTAR checklists	The mean PRISMA score was 19.0 (SD = 4.4) out of a maximum of 27 (71% of items adequately reported, on average). The mean AMSTAR score of 5.2 (SD = 2.9) out of a maximum of 11 (48% of items adequately reported, on average).	The compliance with the PRISMA statement was **moderate** while the compliance with the AMSTAR checklist was **poor**. There were still deficits remaining in the published surgical meta-analyses.
	Agha, R 2016[4] (combined STROBE, CONSORT and PRISMA guideline)	Adherence to three distinct periods for implementation of PRISMA statement: pre, peri, post	Median compliance for period 1 was 48% (ranged from 15%-78%); Median compliance for period 2 was 72% (ranged from 63%-89%); Median compliance for period 3 was 76% (ranged from 48%-96%).	The overall guideline compliance following implementation of the policy **increased** for systematic reviews.
	Aguiar, P 2016[5]	27-item PRISMA and 11-item AMSTAR checklists	The mean (SD) for the PRISMA score was 17.4 (5.6) out of 27. The average total AMSTAR score was 6.9 (2.0) out of 11.	The reporting and methodological quality of systematic reviews and meta-analysis studies on pharmacist interventions in patients with diabetes were **sub-optimal** given the most frequent problems coming from the nonregistration of study protocol, the absence of a list of excluded studies, and the unclear acknowledgment of the conflicts of interests.
	Fleming P, 2013[38]	Modified 27-item PRISMA and 11-item AMSTAR checklists	The mean overall PRISMA score was 64.1% (95% CI 62%–65%).	Reporting of orthodontic systematic reviews in leading orthodontic journals was **deficient** in certain areas, particularly with respect to prospective registration of review protocols, reporting of sources of funding, assessment of reports of risk of bias across studies, definition of summary measures, and detailed

				explanation of the methods of analysis and eligibility criteria.
	Gianola S, 2013[45]	Modified 38-item PRISMA that included the original PRISMA along with 11 additional items	Median=17/27 items (63%); IQR=13-22 948%-82%; Compliance varied across items from 10% to 100%.	The quality of reporting of included RCTs in rehabilitation was **moderate**. Endorsing reporting guidelines can improve poor control of potential source of bias.
	Lee SY, 2016[60]	PRISMA 2009 Statement	The median PRISMA score was 16 of 27 items (59%) (range, 6%-26%; 95% CI, 14%-17%). Compliance varied between PRISMA items. It was poor- est for items related to the use of review protocol (item 5; 4 articles [5%]) and presentation of data on the risk of bias of each study (item 19; 14 articles [18%]). Compliance was highest for description of rationale (item 3; 78 articles [99%]), sources of funding and other support (item 27; 75 articles [95%]), and inclusion of a structured summary in the abstract (item 2; 75 articles [95%]).	The reporting quality of systematic reviews and meta- analyses in plastic surgery needs improvement.
	Li JL, 2014[62]	PRISMA	The range and mean ± SD of overall quality score for included SRs/MAs was 8.5 to 26.0 and 19.6 ± 3.3, respectively; 47 (9.6%) studies had major flaws (an overall score of ≥15), 284 (58.3%) had minor flaws (an overall score of 15.5 to 21.0), and 156 (32%) were considered to have minimal flaws (an overall score of 21.5 to 27.0).	Number of authors

	Liu DN, 2015[64]	PRISMA with a 27 point checklist and AMSTAR with an 11 point checklist	Overall score of 19.9±3.5, but not a single review met all the listing criteria in PRISMA statement. See Table 2 for individual items	Cochrane review, funding resources
	Liu YL, 2014[67]	PRISMA 2009 statement	Among 476 SRs/MAs, only 3 reported the information completely. By contrast, approximately 4.93% (1/203), 8.81% (2/227) and 0.00% (0/46) SRs/MAs reported less than 10 items on the checklist in Chinese journals, international journals, and CDSR, respectively. In general, the least frequently reported items (reported≤50%) in SRs/MAs were item 5 (‘‘protocol and registration’’), 15 and 22 (‘‘risk of bias across studies’’), and 16 and 23 (‘‘additional analyses’’). The remaining items on the checklist were adequately reported (i.e .90%),	not comprehensive
	Ma B, 2012[72]	PRISMA 2009 statement and AMSTAR 2007 statement	Compliance with PRISMA checklist items ranged from 0 to 98.9%.	reporting quality is troubling
	Ma B, 2011[71]	Adherence to PRISMA checklist items	Title, introduction, limitations, and conclusions were reported well in 90% or more of the studies. Most other items varied from 30%–70% of the studies	Compliance with PRISMA reporting guidelines is low for systematic reviews on TCM published in Chinese journals
	Nicolau, I. (2013)[80]	27-item PRISMA	Compliance with PRISMA checklist ranged from 15% to 98%	Overall QOR was **moderate to good**. This may be because most systematic reviews on TB are published in low-impact journals.
	Panic, N. (2013)[83]	27-item PRISMA	Overall PRISMA Compliance by group: Group A: 90.1 (86.4–93.0) Group B: 91.1 (87.6–93.8) Group C: 83.1** (80.0–	Overall QOR **significantly improved** after PRISMA endorsement.

			85.8) Group D: 85.3** (82.4–87.9)	
	Peters, J.P.M. (2015)[88]	PRISMA 2009	Overall PRISMA compliance by journal: In the top 5 Ear Nose Throat journals reported a median of 54.4% (mean 62.2%, 95% CI: 54.4%-71.7%); In the 'gold standard' Cochrane Database of Systematic Reviews reported a median of 100.0% (mean 98.2%, 97.3%-99.1%) Overall PRISMA for abstracts items compliance by journal: In Ear Nose Throat journals reported a median of 41.7% (mean 31.7%, 30.2%-44.0%); In the 'gold standard' Cochrane Database of Systematic Reviews reported a median of 75.0% (mean 75.2%, 73.1%-76.2%)	Overall QOR was **suboptimal**.
	Rice, D.B. (2016)[93]	Adapted PRISMA	Mean for pre-PRISMA was13; Mean for post-PRISMA was 17.	Overall QOR had **deficiencies** both before and after publication of PRISMA.
	Tan, WK (2014)[103]	PRISMA	The average percentage of PRISMA items reported in both general and vascular surgical journals in 2012 was 73%, compared to 65% in 2008, indicating some improvement in the quality of reporting ($p < 0.01$), although this is clearly way short of ideal.	Substandard
	Tunis, A (2013)[105]	PRISMA	Average PRISMA was 21.8 of 27	Modest
	Willis, B 2011[115]	Adherence to the 27 items PRISMA	Of the 236 meta-analyses included following	Compliance with the PRISMA statement was

		checklist.	selection: 1% reported the study protocol; 25% reported the searches used; 32% reported the results of a risk of bias assessment; and 35% reported the abstract as a structured summary.	**generally poor**: none of the review completely adhered to all 27 checklist item for published meta- analyses of diagnostic tests.
	Weir, C 2012[113] (combined PRISMA and QUOROM guideline)	A integrated score consisting of the number of items completed over the total numbers of items on both the PRISMA and QUOROM criteria, resulting in cored ranking from 0 to 100% (excluding the items focused on the abstract)	Mean = 63% (range 45-81%) on a scale of 0-100%	Systematic reviews of empirical computerized provider order entry research had **moderate quality**.
QUOROM	Al Faleh, K 2009	18-item QUOROM checklist, 10-item checklist OQAQ used for scientific quality	Cochrane Neonatal review Group systematic reviews scored a mean of 4.5 (0.9), (95% CI of 4.27–4.77); Mean (SD) of OQAQ total scores was 4.5 (0.9) (95% CI 4.27–4.77).	The quality of reporting of systematic reviews published in Cochrane Neonatal review group were **good with minor flaws**.
	Bereza, B 2008[15]	18-item QUOROM checklist, 10-item checklist OQAQ used for scientific quality.	61±19% (median 60%, range 39-94%) for QUOROM checklist. 58% ± 28% for OQAQ	"Reporting/scientific quality was considered **less than fair-to-good.** Stakeholders should strive for higher scientific quality of meta-analyses".
	Biondi-Zoccai, G 2006[17]	18-item QUOROM checklist	Median compliance with the QUOROM checklist was 16 (range 11 to 17).	Overall compliance with the QUOROM checklist was **relatively good**.
	Hemels M, 2004[49]	18-item QUOROM checklist	On average 50.2% of the QUOROM items were reported.	The overall quality of reporting in meta-analysis of RCTs in major depressive disorder was marginally acceptable.
	Junhua Z, 2007[53]	28-item QUOROM checklist, ten-item checklist OQAQ used for scientific quality	No combined score; methodological and reporting flaws in more than half of the review articles. Flaws were mainly in the literature search,	Methodology and reporting quality are poor in both systematic reviews and meta-analysis of TCM published in journals in the People's Republic of China

			characteristics of included and excluded studies, quality assessment of primary trials, and data merging	
	Shea, B 2006[99]	18-item QUOROM checklist, 10-item checklist OQAQ used for scientific quality	N/A for overall mean difference of QUOROM; Mean difference of OQAQ was 0.11 (-0.28; 0.70 p = 0.52)).	The overall quality of Cochrane reviews was **fair-to-good**. No overall improvement seen with updating and methodological quality even though quality of reporting improved on certain individual items.
	Shea,B 2006 [100]	18-item QUOROM checklist, 10-item checklist OQAQ used for scientific quality.	All systematic reviews were found to have good overall quality. OQAQ mean score was 5.02 (95% CI 3.71-6.32).	Reporting quality of Cochrane musculoskeletal systematic reviews was generally good, with room for improvement.
	Vigna-Talianti, F 2006[107]	QUOROM-based checklist (score out of 50)	29.9/50	"Oncologists should be aware that they could be relying on poor underlying documents. Writing groups should be aware of methodological problems, and should consult the existing manuals for the preparation of guidelines".
	Weir, C 2012 [113](combined PRISMA and QUOROM guideline)	A integrated score consisting of the number of items completed over the total numbers of items on both the PRISMA and QUOROM criteria, resulting in cored ranking from 0 to 100% (excluding the items focused on the abstract)	Mean = 63% (range 45-81%) on a scale of 0-100%	Systematic reviews of empirical computerized provider order entry research had **moderate quality**.
	Wen, J. (2008)[114]	18-item QUOROM	The mean overall QUOROM score of the 161 articles was 12.3 (95% CI: 12.0, 12.6) with minimum=4.5 and maximum =16.5.	Overall QOR were **basically acceptable**, but **in need of improvements**, which can be seen overtime.
STROBE	Agha, R 2016[3]	22-item STROBE checklist	The average (SD) STROBE score was 12.4 (3.36)	The reporting quality of observational studies in

			(56%; range, 2–20.1).	Plastic Surgery needed **improvement**.
	Agha, R 2016[4] (combined STROBE, CONSORT and PRISMA guideline)	Adherence to three distinct periods for implementation of STROBE statement: pre, peri, post	Median compliance for period 1 was 68.4% (ranged from 44%-88%); Median compliance for period 2 was 67.2% (ranged from 46%-84%); Median compliance for period 3 was 76.5% (ranged from 54%-97%).	The overall guideline compliance following implementation of the policy **increased** for observational studies.
	Cook, D 2011[26]	Quality of reporting/methodological quality and association between methodological quality and effect size	253 (90)	Reporting quality of health professions education experimental research was found to be generally **suboptimal**.
	Delaney M, 2010[29] (combined STROBE and CONSORT guideline)	Adherence to CONSORT and STROBE checklists	Observational studies (STROBE): 65.2% did not describe the magnitude and direction of potential bias; 63.2% did not make clear adjusted confounders and why if they were included; 63.2% showed unaddressed missing data; 57.3% did report confounder-adjusted and the precision of estimates.	Both observational studies and RCTs **missed reporting some important factors**. There were gaps in methodologic reporting. Using CONSORT and STROBE checklists may improve the **deficiencies** in reporting.
	Fung A, 2009[42] (combined STROBE and CONSORT guideline)	STROBE (Max Score of 37 points) statement indicators	STROBE mean and median: 70% and 71%, respectively	Overall level of reporting is acceptable, and has improved since the creation of CONSORT and STROBE.
	Parsons, N.R. (2011)[84]	Overall compliance calculated as the weighted mean of the compliance rates for the seven selected journals, using a previously made questionnaire	N/A	Very few papers fulfilling all criteria; general lack of statistical rigor
	Rao, A.	STROBE 2007	Pre- and post- STROBE	Overall QOR had

	(2016)[92]		period analyses revealed an increase in manuscript STROBE score (median score 77.8% (IQR, 64.7–82.0%) vs 83% (IQR, 78.4–84.9%), p = 0.04)	**deficiencies** both before and after publication of STROBE.
	Shawyer, A.C. (2015)[98]	22-item STROBE	The median quality of reporting across all included studies was 65% (minimum=48%, maximum=95%)	Overall QOR was **moderate**, with a trend of **improvement** overtime.
STARD	Fidalgo B, 2015[37]	14-item QUADAS and 25-item STARD checklists	The median of QUADAS items was 9 out of 14 (IQR 7-10); The median score of STARD was 11 out of 25 (IQR 10-14).	The quality of reporting and methodology of diagnostic accuracy studies of perimetry was **sub-optimal**. **No substantial improvement** was showed after the development of the STARD guideline.
	Fontela P, 2009[39]	14-item QUADAS and 25-item STARD checklists	The percentages for the score of QUADAS varied from 6% to 98%; The percentages for the score of STARD varied from 0% to 99%.	The quality of reporting of recently published diagnostic accuracy studies on commercial tests for tuberculosis, malaria and human immunodeficiency virus was **moderate to low** and **poorly reported**. Endorsing QUADAS and STARD can improve the methodological and reporting quality for diagnostic accuracy studies in infectious diseases.
	Freeman K, 2009[40]	25-item STARD checklists	Scores ranged from 5 to 13 out of 25.	The quality of reporting for studies reporting diagnostic accuracy was **poor**. Adherence to STARD checklist was recommended especially for the reporting of the diagnostic accuracy of non-invasive prenatal diagnostic test.
	Miller E, 2009[75]	STARD (Standards for Reporting of Diagnostic Accuracy 25 items)	Not reported	Fair
	Zafar, A.	STARD	The mean score (SD) was	Overall QOR was

	(2008)[117]		19.8 (6.5) out of a maximum of 50.	**suboptimal**.
	Zintzaras E, 2012[123]	Percentages of 22 methodological related items of the STARD checklist for each item	Percentage varied from 7.8% to 100.0%.	The overall reporting quality for studies of diagnostic accuracy of anti-CCP was **relatively good** with some **essential methodological issues**. The reporting quality needed further improvement.
ARRIVE	Bramhall, M 2015[20]	Adherence to ARRIVE checklist	The mean weighted score across all colitis models was 81.7% (SD = +/- 67.038).	The quality of methods reporting in modeling colitis had serious flaws. Endorsing the ARRIVE checklist can improve the quality of publications in this field.
	Gulin J, 2015[47]	Adherence to pre- and post-ARRIVE checklist	N/A	A significant **lack of compliance** with ARRIVE guidelines in research involving animals for testing of efficacy of new compounds for Chagas disease treatment was observed. Endorsing ARRIVE guideline was **not sufficient to improve** reporting of animal studies.
	Liu YL, 2016[68]	ARRIVE (Animal Research: Reporting of In Vivo Experiments) 2010 Statement	The range of ARRIVE score is from 12 to 27 with a maximum possible score of 40. The value for each of the median (P25, P45) ARRIVE checklist scores for studies published during January and June2010, July and December 2010, 2011, and 2012 was18.50(17.00,20.00), 19.00 (18.00,21.00), 19.00 (18.00,21.00) and 20.00 (18.00,22.00), respectively. Studies published in 2012($P = 0.012$), 2011 ($P = 0.015$), 2010, July~Dec ($P<0.017$) had a significantly larger ARRIVE checklist score than those published in Jan.~June,2010,respectively.	Poor

	Schwarz, F. (2012)[96]	ARRIVE 2010	N/A	Overall QOR was **variable** across categories.
MOOSE	Zhang Z, 2015[118]	Adherence to MOOSE 2010 and AMSTAR 2010 checklists	Compliance with the MOOSE checklist items ranged from 0% to 96.7%; Compliance with the AMSTAR checklist items ranged from 4.5 to 75.8.	The reporting quality of meta-analyses of observational studies that have been published in Chinese journals was **questionable**.
CHEERS	Aguiar, P 2016[6]	Adherence to 24-item CHEERS checklist	The average score was 14.6 (SD = 2.6) (ranged from 11–18).	The reporting quality of the economic studies of novel therapeutic agents in multiple myeloma was **poor**. Improvement was needed.

Abbreviations: **SD**, standard deviation; **CONSORT**, Consolidated Standards of Reporting Trials; **RCT**, randomized controlled trial; **CHM**, Chinese herbal medicine; **HAART**, highly active antiretroviral treatment; **CLEAR NPT**, checklist to evaluate a report of a nonpharmacological trial; **CABG**, coronary artery bypass surgery; **STROBE**, Strengthening the Reporting of Observational Studies in Epidemiology; **NMECI**, Nelson-Moberg Expanded **CONSORT** Consolidated Standards of Reporting Trials Instrument; **PedCAM**, Pediatric Complementary and Alternative Medicine Research and Education Network; **TOS**, thoracic outlet syndrome; **N/A**, not applicable; **NMENCI**, Nelson-Moberg, Norton Expanded Consolidated Standards of Reporting Trials Instrument; **CI**, confidence interval; **URM**, uniform requirement for manuscript; **CRT**, consolidated standard of reporting trial; **NI**, noninferiority; **PRISMA**, preferred reporting items for systematic reviews and meta-analyses; **QUOROM**, QUality of Reporting of Meta-analysis; **OQAQ**, Overview Quality Assessment Questionnaire; **TCM**, traditional Chinese medicine; **JAAD**, the Journal of the American Academy of Dermatology; **BJD**, the British Journal of Dermatology; **NPT**, non-pharmacological trials; **IQR**, interquartile range; **STRICTA**, Standards for Reporting Interventions in Clinical Trials of Acupuncture; **JCIM**, Journal of Chinese Integrative Medicine; **CJIM**, Chinese journal of Integrative Medicine; **OQS**, overall quality score; **PRO**, Patient Reported Outcomes; **QUADAS**, Quality Assessment tool for Diagnostic Accuracy Studies; **GSPC**, Gold Standard Publication Checklist.